This memorandum provides a review of several dispute resolution issues outlined in the Newtown Creek Group's (NCG) letter to EPA Region 2 (NCG letter, 2017) for the City of New York (the City). Specific issues addressed here include:

1. The identification of confounding factors in the NCG development of correlations of toxicity vs pore water chemistry;
2. The physical effects of oil should be considered when interpreting toxicity test results;
3. The use of ten day toxicity testing in the assessment of benthic community risks;
4. The use of No Observed Adverse Effects Levels (NOAELs) in the risk assessment; and
5. The use of reference area data in the ecological risk assessment.
6. The estimation of BSAFs should follow Burkhardt's recommendations

1. THE IDENTIFICATION OF CONFOUNDING FACTORS IN THE NCG DEVELOPMENT OF CORRELATIONS OF TOXICITY VS PORE WATER CHEMISTRY

In recent technical discussions of the disputed areas, EPA has written that if NCG includes *"a robust discussion about other possible reasons for the toxicity (including but not limited to, bulk sediment comparisons, concentrations of individual compounds and DNAPL), the discussion and figures that were identified as needing to be deleted can remain in the document."* Adding additional robust discussion (as requested by EPA) <u>does not</u> justify the continued inclusion of the flawed NCG analysis in the Baseline Ecological Risk Assessment (BERA).

The NCG evaluated the relationship between a compound parameter, the sum of PAH TUs plus the sum of metal TUs, and toxicity test results using a selected set of triad data from the BERA field program. This evaluation has several large sources of uncertainty in the selection process and the approach, which should disqualify this evaluation from consideration in the BERA.

The NCG evaluation:

A. Selected a subset of triad stations for the analysis based on two highly uncertain and insufficiently supported criteria:
   i. Elevated C19 to C36 aliphatic hydrocarbons greater than the Stanley et al. (2010) mineral oil benchmark; and
   ii. TU less than 2 for porewater PAH (34) or SEM metals.

B. Attempted to correlate an unsupported compound parameter (sum of PAH TUs plus sum of metal TUs); and

C. Attempted to seek a correlation between metals and toxicity when in fact, the lines of evidence in the BERA indicate that metals are not likely to be toxic.

A. <u>The NCG selected a subset of triad stations for the analysis based on two highly uncertain and insufficiently supported criteria</u>.

In the BERA, the NCG selected a subset of triad stations for regression analysis based on two highly uncertain and insufficiently supported criteria: (1) elevated C19 to C36 aliphatic hydrocarbons greater than the Stanley et al. (2010) mineral oil benchmark and (2) TU less than 2 for porewater PAH (34) or SEM metals. On the basis of these two criteria, NCG eliminated nine stations (seven Newtown Creek stations and two reference area stations) from their analysis in an attempt to assign the cause of toxicity to the proximity of sample locations to CSOs or stormwater discharges. This was modified in the February, 2017 NCG summary memorandum to EPA (page 9), the rationale for the selection of these nine stations was characterized as *"stations for which the toxicity test results are not consistent with expected pore-water based concentration-response relationships"* and the results at these stations were explained by their spatial proximity to CSOs and municipal outfalls. Table 8-9 of that memorandum indicates that these stations were *"removed as confounding factors due to C19-C36 concentrations"*. The City comments that follow assume that the reasoning provided in the BERA is the operating rationale for station removal as the current dispute indicates that reasoning may stay as long as other lines of evidence are presented.

A.i. Mineral Oil Benchmark Should Not Be Used to Screen Triad Stations against Concentrations of C19 to C36 Aliphatics.

The NCG bases the first criteria, the mineral oil benchmark, on a long chain of weakly linked assumptions starting with the potential toxicity of Unresolved Complex Mixtures (UCMs), the measured toxicity of mineral oil, and a weakly supported argument that the mineral oil benchmark is reflective of toxicity from Extractable Petroleum Hydrocarbon (EPH) fraction C19 to C36. The assumptions required to develop this argument are too uncertain to be included in the risk assessment section of the BERA. The NCG attempts to make the case that hydrocarbon UCMs may be confounding toxicity in Newtown Creek sediments based on a chain of assumptions that vaguely implicate CSOs as the source of this confounding factor. The BERA uses the following chain of assumptions: (1) UCMs have been shown to be toxic to benthic organisms elsewhere (this is true of all COPCs); (2) saturated hydrocarbon (oil) has been shown to be as much as 90% UCM; (3) EPH was measured in Newtown Creek triad sediments; (4) EPH includes an aliphatic hydrocarbon range; (5) the literature offers an experiment that provided a LC-50 from a 10-day *Leptocheirus* test using mineral oil in which a concentration of 210 mg/kg elicited an effect, assumed to be physical; (6) mineral oils have carbon ranges of C15 to C50 and the boiling point of mineral oil, C19 alkane and C32 alkane are similar; (7) therefore, the 210 mg/Kg LC-50 for mineral oil is applied as a sediment benchmark for the C19 to C36 EPH fraction measured in Newtown Creek, which is assumed to be a good surrogate for mineral oil.

There are a number of flaws in this chain of logic that invalidate the development of a sediment benchmark for EPH including: (1) a lack of explanation about how specifically the comparison of alkane boiling points to mineral oil boiling points supports the toxicological extension to C19 to C36 EPH fractions; (2) the BERA's assumption that EPH is a reasonable surrogate for mineral oil based on the range of carbon numbers is not supported by Mount et al., 2010, who state that mineral oil is generally in the range C13 to C24 rather than the higher range in EPH C19 to C36 fraction; (3)The BERA ignores the range of LC-50s for mineral oil provided in their cited reference (Stanley et al., 2010) which indicates that the LC-50 ranges from 110 to 210 depending on the beaker size and number of test organisms.

In particular, the BERA applies no uncertainty factors, as is the standard of practice to the development of a benchmark, despite the various clear sources of uncertainty such as: (1)the BERA inappropriately uses the highest LC-50 reported for mineral oil (Stanley et al. 2010 also report a LOEC of 0.15 mg/kg); (2) the assumption that the C19 to C36 fraction of EPH is a surrogate for UCMs, which is a surrogate for petroleum products, a broad mixture; (3) as indicated above, there is a range of possible benchmarks ranging from 0.15 (LOEC) to 210 mg/kg (EC-50); and (4) these ranges of effect levels were derived from a 10-day exposure and thus may overestimate the exposures associated with more chronic exposures.

Further, the NCG has not proven that the elevated C19-C36 is due to CSOs or MS4s. No data has been presented to support attribution of elevated C19 to C36 fraction to CSOs and MS4s, and without the measurement of C19 to C36 compounds in the discharge, there is no basis to assign C19 to C36 compound contamination detected in the sediments to any point source discharges. Data is available at some upland sites, which shows C19 to C36 compound concentrations at high concentrations. For example, the C19 to C36 concentration in the soils at the upland site Quanta (former refinery), are elevated, with an average concentration of 480,000 mg/kg (nearly 50 percent). Without available data from all sources (upland Sites, NAPLs, CSOs and MS4s) the assertion by the NCG is arbitrary and needs to be deleted.

This uncertain benchmark should not be applied as a criterion to remove stations from an analysis of porewater chemistry vs toxicity.

A.ii. The Sum PAH TU <2 is an Inappropriate Screening Criteria for Triad Stations

The NCG uses a second selection criteria, TU of less than 2 for porewater PAH (34) or SEM metals to select triad stations to eliminate from their evaluation. The rationale for this criterion is that stations with a TU less than 2 for either of these parameters will select stations that are not predicted to be toxic due to exposure to either PAHs or metals. The application of this criterion results in the elimination of seven Newtown Creek stations. The criterion, however, misuses the EPA thresholds for predicting the likelihood of toxicity. That threshold TU specified by EPA for either SEM metals or PAHs is 1, not 2. Furthermore, EPA specifies the threshold as a categorical threshold, not a continuous variable. Specifically, EPA (Burgess et al., 2013) explicitly state that *"For the interstitial water approach. . .when the metal mixture interstitial water ESB >1, sediment toxicity due to metal mixtures may occur, while in cases where the ESB value is ≤ 1, toxicity due to metals is unlikely."* Similarly for PAHs, EPA (2003) states that *"Benthic organisms should be acceptably protected from the narcotic effect of PAH mixtures …if the ΣESBTU is less than or equal to 1.0 and if the ΣESBTU is greater than 1, sensitive benthic organisms may be adversely affected"* by direct toxicity. In both instances, the threshold is 1 rather than the value of 2 used in the NCG selection process. This unsupported inflation of the well documented EPA threshold results in the elimination of three stations in which the PAH TU is greater than 1. Using the correct threshold (1), these sediments are likely to be toxic, according to EPA methodology. The NCG is claiming these stations as having sediments that are not toxic due to PAH exposure when the EPA guidance explicitly states that they may be adversely affected, and in fact, these stations exhibited sediment toxicity consistent with the EPA prediction.

The application of this criterion allowed NCG to screen out three site stations from their analysis that had sum PAH TUs > 1, which indicates that these stations are likely to be toxic. NCG's raising the threshold to a value of 2 is not supported by EPA guidance regarding the application of the sum PAH ESB and results in an arbitrary screening of data from the analysis.

B. <u>NCG attempted to correlate an unsupported compound parameter (sum of PAH TUs plus sum of metal TUs) with toxicity.</u>

NCG provides no technical support for adding two independent parameters as one compound parameter in the evaluation of confounding factors. There is no toxicological reason to add these parameters. EPA guidance (EPA, 2003; Burgess, 2013) justify the sum PAH TU as an indication of whether a sediment sample may or may not be toxic based on the supported assumption that the individual PAHs in that summation are all acting with the same toxic mechanism, narcosis. EPA's use of the sum metals TU does not rest on the same assumption that the toxicological mechanism for the metals is narcosis. The metals may all have different modes of action, none of which EPA assumes are narcosis. NCG provides no evidence that there is any toxicological justification for adding these completely different and differently derived summations.

In addition, as described above, these parameters, sum AVS-SEM TU and sum PAH TU, are categorical in that they are interpreted based on a threshold. NCG has used these parameters as a continuous variable in a correlation without supporting the use in this manner.

C. <u>NCG attempts to seek a correlation between metals and toxicity when in fact, the lines of evidence in the BERA indicate that metals are not likely to be toxic.</u>

EPA (EPA, 2005; Burgess, 2013) explicitly recognize three lines of evidence that address whether sediment metals (the SEM metals) are likely to be bioavailable in their dissolved forms in pore water and therefore likely to be toxic.

The data in the BERA clearly demonstrate that the SEM metals are not a likely cause of toxicity in any of the sediment samples based on the EPA ESB methods and interpretive framework. Specifically, EPA (2005) states that "benthic organisms are sufficiently protected if the sediment meets either one of the following benchmarks":

(1) Σi [SEMi] $\leq$ [AVS]

or

(2) Σi [(Mi,d)/(FCVi,d)] $\leq$ 1.0 (for the five SEM metals)

In addition, EPA (2005) uses a third approach to refine the uncertainty associated with the benchmark:

(3) (ΣSEM-AVS)foc

EPA uses this TOC corrected SEM-AVS approach (item 3 above) to refine the uncertainty associated with the benchmarks and recognizes three interpretive levels, one of which is that if the (ΣSEM-AVS)foc < 130, then toxic effects are not expected.

The data in the BERA clearly indicates that metals are unlikely to be the cause of benthic toxicity based on the analysis methods presented above, EPA (2005). Specifically, among the 60 triad stations that NCG used in their analysis, every station met both conditions 1 and 3, indicating with considerable certainty that the benthic organisms are sufficiently protected from exposure to SEM metals in pore water. In addition, 44 stations met condition 2.

Using the interpretive methods from EPA (2005) and Burgess (2013) these data indicate that the benthic community is not at risk from exposure to SEM metals through direct contact at any of the 60 triad stations, because at least one of the first two criteria above are met and criteria 3 provides an added level of certainty that the benthic community is not at risk from direct exposure to SEM metals.

In the BERA, NCG applies the SEM-AVS benchmarks to "bulk sediment." However, the derivation of the SEM-AVS model by EPA (2005) clearly states that "*partitioning models can relate sediment* *concentrations for cationic divalent metals (and monovalent silver) on an AVS basis to the absence of freely-dissolved concentrations in interstitial water".*

SUMMARY - THE IDENTIFICATION OF CONFOUNDING FACTORS IN THE NCG DEVELOPMENT OF CORRELATIONS OF TOXICITY VS PORE WATER CHEMISTRY

The technical objections to the NCG approach presented in this Section 1 demonstrate the lack of a valid technical justification for: (1) the C19 to C36 screening criteria that NCG used to select stations for the correlation analysis; (2) the arbitrary selection of a toxicity unit threshold of two for the second screening criteria; and (3) the unsupported and toxicologically meaningless parameter that combines PAH and metal toxicity (sum PAH TU plus sum SEM TU). This lack of a technical justification for these NCG analyses is sufficient reason for these analyses to be removed from the BERA. In recent discussions between EPA and NCG, EPA risk assessors have indicated that these analyses are not compelling and do not demonstrate any relationship between CSOs and observed toxicity.

Inclusion of additional lines of evidence analysis requested by the EPA does not validate NCG's flawed reasoning on confounding factors and toxicity. This should be removed from the BERA. In addition, any new reasoning should be reviewed by all stakeholders before acceptance. Based on the deficient NCG submission it would be more effective if the EPA or the City developed the approach in the BERA, rather than have NCG submit another deficient analysis that would require further discussion, revision and possible subsequent resubmission.

2. THE PHYSICAL EFFECTS OF OIL SHOULD BE CONSIDERED IN INTERPRETING TOXICITY TESTS

The analysis of sediment toxicity and the evaluation of the source of toxicity in Newtown Creek should recognize that the City's measurement of sediment toxicity throughout Newtown Creek demonstrates that there are likely two populations of sediment samples based on clear differences in the visible presence of oil in the toxicity test samples (presence or absence), the higher concentrations of Total Petroleum Hydrocarbons (TPH) in those samples with visible presence of oil, obvious differences in toxicity (10- and 28-day survival), and the location of these stations in the upper reaches of Newtown Creek (Turning Basin and tributaries).

Figure 1 shows the locations of sediment toxicity test samples where the laboratory (USACE ERDC lab Vicksburg, MS) observed evidence (sheens, NAPL) of separate phase oil in test samples. Most of the samples in the upper reaches of the Newtown Creek exhibited visible evidence of oil contamination. Concentrations of TPH in these upper reach stations were compared to the TPH concentrations in stations in lower Newtown Creek. Figure 2 shows that these two groups (lower Newtown and upper reach stations) have notably different concentrations of TPH.

A comparison of the 10-day toxicity (Figure 3) and the 28-day toxicity (Figure 4) show that these two groups have very different toxicological responses. These differences may be due to the physical effects that oil has on the respiratory systems of marine invertebrates. These data suggest, but do not test this hypothesis (if the hypothesis is correct, then it would be futile to seek a chemical cause for toxicity in those stations where the physical effects of oil are killing the organisms before any chemical effects can be realized). However, it is clear from these figures that there appears to be a bimodal and discontinuous distribution of toxicity in the samples tested by the City and that this bimodal distribution can be described by station differences in observations of oil, concentrations of TPH, and location in the upper reaches of the creek. This distribution calls into question the validity of attempts to assign singular and similar sources of toxicity to the pooled group of stations in Newtown Creek.

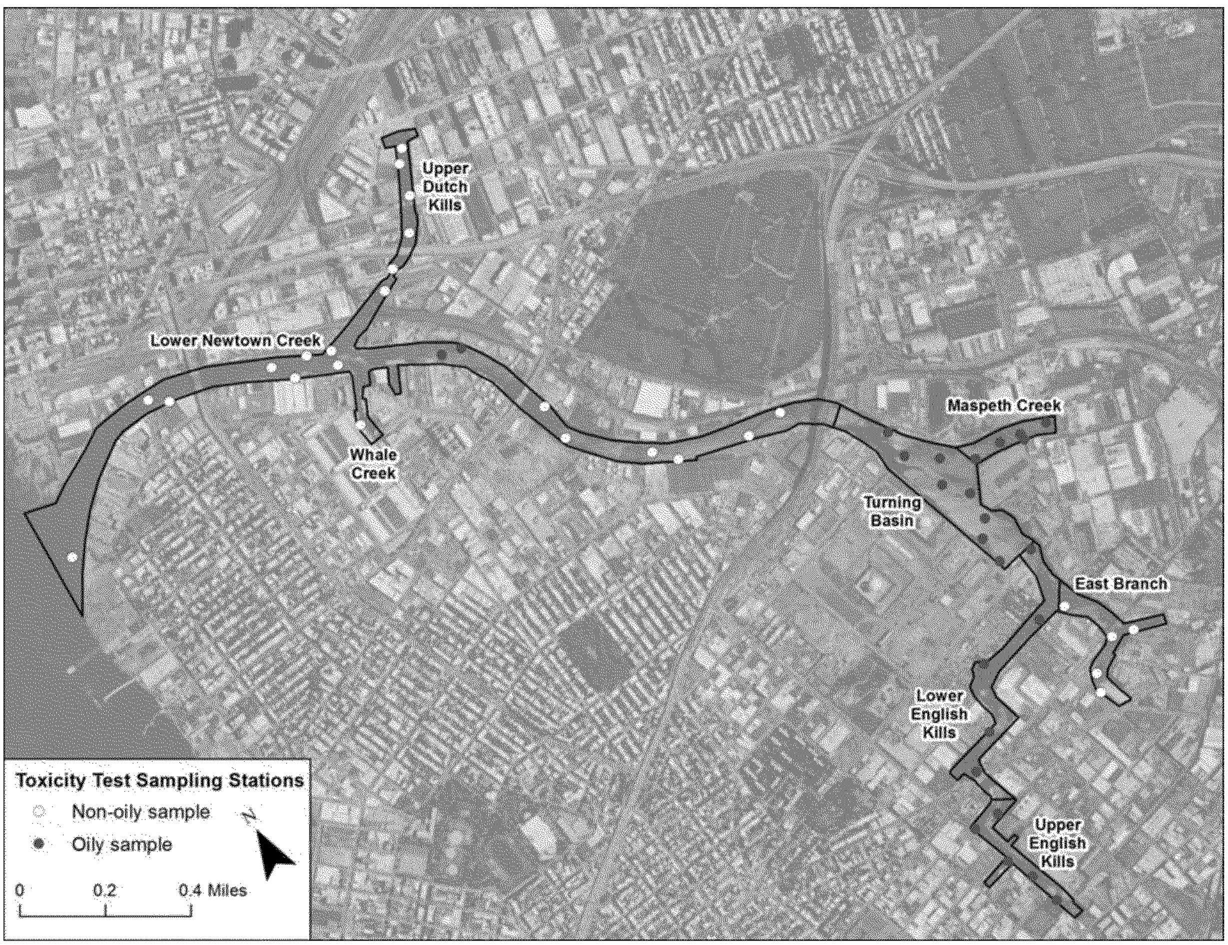


Figure 1. Visual observations of oil in the toxicity test sediment samples

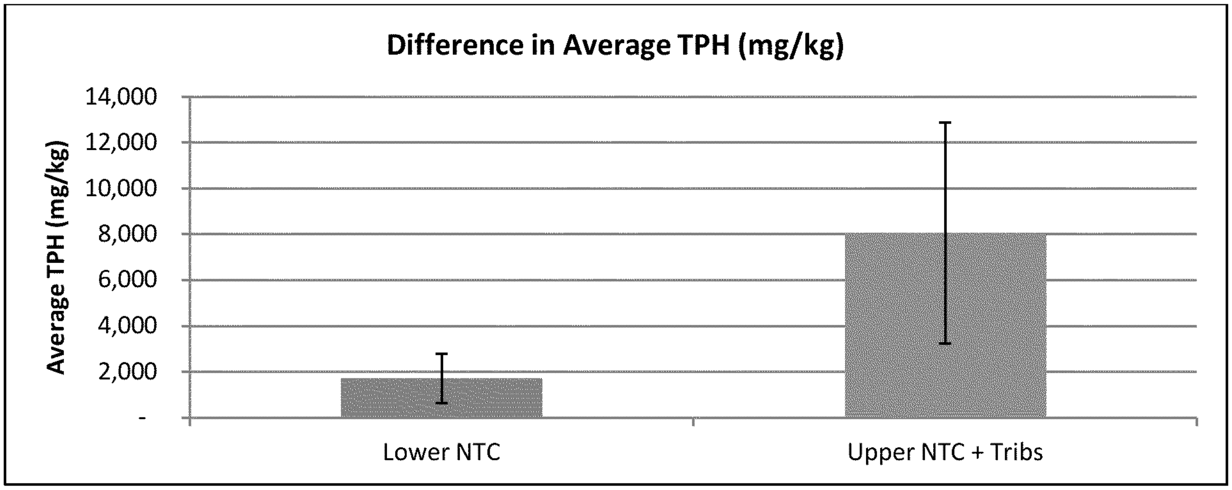


Figure 2. Average TPH (mg/kg) concentrations in Lower Newtown Creek (blue) and Upper Newtown Creek and tributaries (green). Error bars represent one standard deviation

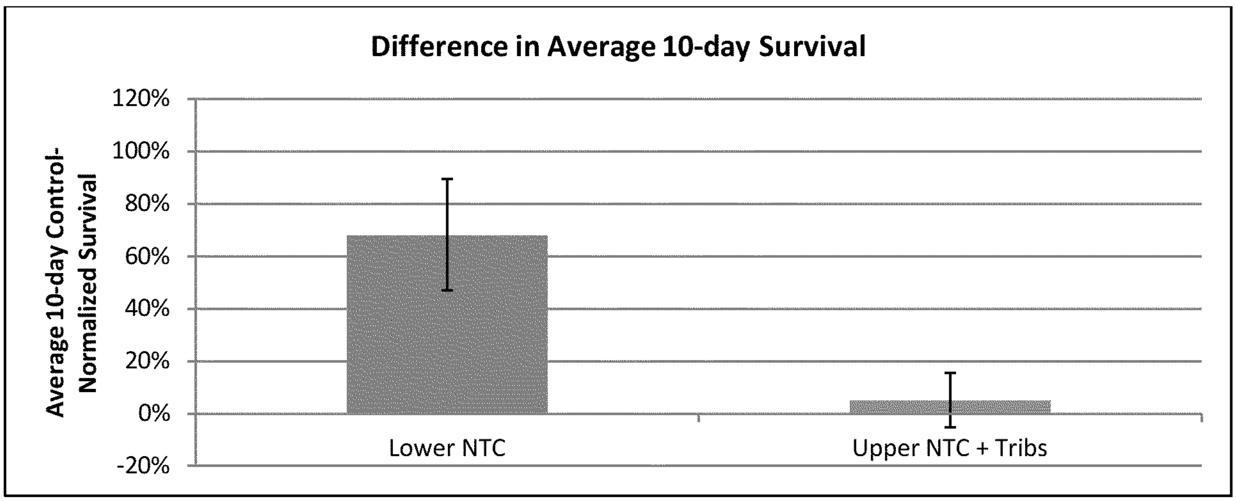


Figure 3. Comparison of 10-day survival results in Lower Newtown Creek (blue) and Upper Newtown Creek and tributaries (green). Error bars represent one standard deviation

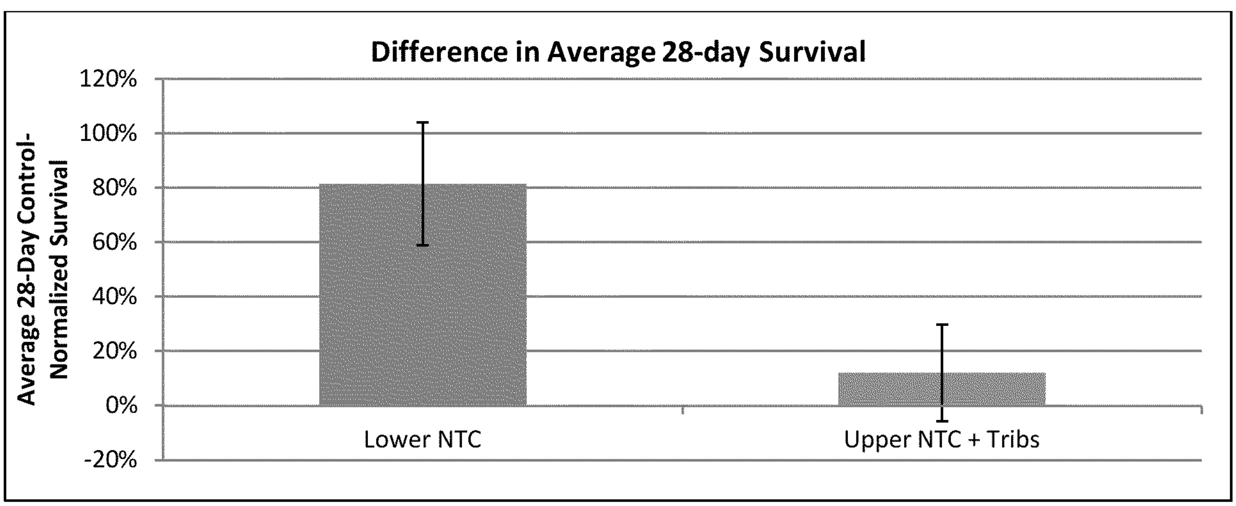


Figure 4. Comparison of 28-day survival results in Lower Newtown Creek (blue) and Upper Newtown Creek and tributaries (green). Error bars represent one standard deviation.

3. THE USE OF TEN DAY TOXICITY TESTING IN THE ASSESSMENT OF BENTHIC COMMUNITY RISKS

The toxicity testing shows that:

- The 10-day and 28-day tests were conducted according to standard methods, and met their respective performance standards. There is no compromise or bias in either of these tests;
- These tests should be interpreted in light of their different purposes and methods for assessing acute toxicity in the case of the 10-day test and chronic (including sub-lethal endpoints) in the case of the 28-day test;
- The NCG argument regarding the variability of one test over the other is simply a misreading of the literature that they cite regarding this topic.

NCG incorrectly characterized the 10-day toxicity tests as "compromised", noting the feeding and water change differences between the 10- and 28-day toxicity test protocols. Both the 10- and 28-day toxicity tests conducted for Newtown Creek followed standard, approved protocols, and met all required conditions throughout the tests. The controls had acceptable survival in the 10-day tests, indicating that test conditions did not compromise the testing. Therefore, the results of both tests are equally valid for their individual purposes.

Various federal agencies recognize that these two tests are separate and independent measures of either acute or chronic toxicity (EPA, 2014; EPA, 1997; USACE and EPA Region 2, 2016). As such, one cannot be characterized as biased in comparison to the other. They are measuring different properties. The discrepancy between 10- and 28-day survival results is not due to a "compromised" 10-day test, but rather to the inherent differences between the two tests. As NCG pointed out, the tests differ in feeding and water change regimes. They also differ in light regime. In a 10-day toxicity test, *L. plumulosus* is kept under a 24-hour light regime. Since the organisms innately avoid light, this effectively drives the organisms into the sediment for the entire duration of the test. Constant immersion in the sediment allows *L. plumulosus* to act as surrogates for burrowing benthic macroinvertebrates that are in constant contact with the sediment. In a 28-day test, the light regime is adjusted to a more natural cycle (16 hours of light followed by 8 hours of darkness each day). In the absence of direct light, *L. plumulosus* are more likely to exit the sediment and swim in the overlying water. This change in conditions is less about providing a more hospitable environment for the organisms, and more about providing conditions in which the organisms will mate (a necessary precursor for measuring reproduction as an endpoint). These are different lines of evidence, each with its own separately developed methodology for different purposes. The 10-day test is designed as an indication of acute toxicity, while the 28-day test is designed as an indicator of sublethal toxicity.

NCG, in their response to EPA comments, cites a paper (Kennedy et al. 2009), claiming that it demonstrates the variability of the 10-day test, and NCG states that *"in an ecological risk assessment, a 10-day test measuring acute effect is not as strong of a line of evidence as a 28-day test measuring chronic endpoints"*. However, Kennedy et al. (2009) actually demonstrate the opposite in terms of variability. They note that the *"10-d A. abdita, 10-d L. plumulosus and 28-day L. plumulosus tests were comparable between laboratories,"* but note that *"intra-treatment sub-lethal endpoint variability was greater"* and *"chronic L. plumulosus test method was less consistent among laboratories relative to acute test methods"* and the authors demonstrate that the 28-day sub-lethal endpoints may be either more or less sensitive than the 10-day acute test in identifying toxicity. The results of the Kennedy et al. paper do not support the NCG statements regarding variability, or bias of the 10-day test.

Subsequently in their recent (March 10, 2017) summary letter regarding the dispute resolutions, NCG cites EPA, 1994 and EPA 2014 to support the position that the chronic tests are more appropriate. The EPA, 2014 is a memo from EPA office of pesticides that addresses the testing of a single chemical (new pesticide registrations) and recommends the use of subchronic tests (10-day) when new pesticide half-lives are short, and chronic tests (28-day) when new pesticide half-lives are longer. The cited guidance is not appropriate for a mixed chemical testing that occurs at a Superfund site. The EPA, 1994 guidance does not address 10-day vs 28-day tests but sets some general recommendations regarding the use of chronic and acute tests, which EPA defines as 24 to 96 hour tests (much less than the 10-day test used at the Newtown Creek site).

The City is in agreement with EPA that the 10-day toxicity test is a standard, well-documented, and unbiased toxicity test and is valid as a separate, independent, and equally weighted line of evidence for assessing risk to benthic invertebrates. As such, the ten day test carries as much weight as the 28 day toxicity test.

4. THE USE OF NOAELS IN THE ASSESSMENT OF ECOLOGICAL RISK

The BERA uses both NOAELs and Lowest Observed Adverse Effects Level (LOAELs). NOAELs are applied in the Phase II screening process (Section 5). LOAELs are applied in Wildlife Risk Assessment in Section 11. The application of NOAELs in the risk screening is appropriate, the Risk Assessment Guidance for Superfund (EPA 1997) is clear that both a NOAEL and LOAELs are needed to bound the wildlife risk estimates. EPA (1997) emphasizes how these effects values should be included and states:

> *Section 7.3.1: "Key outputs of the risk characterization step are contaminant concentrations in each environmental medium that bound the threshold for estimated adverse ecological effects given the uncertainty inherent in the data and models used. The lower bound of the threshold would be based on consistent conservative assumptions and*

NOAEL toxicity values. The upper bound would be based on observed impacts or predictions that ecological impacts could occur. This upper bound would be developed using consistent assumptions, site-specific data, LOAEL toxicity values, or an impact evaluation."

Additionally, EPA (1997) discusses that the threshold for potential effects is a range between the no effect level and the lowest effect level. The guidance states (EPA 1997),

Section 7.5: "Risk characterization integrates the results of the exposure profile and exposure-response analyses, and is the final phase of the risk assessment process. It consists of risk estimation and risk description, which together provide information to help judge the ecological significance of risk estimates in the absence of remedial activities. The risk description also identifies a threshold for effects on the assessment endpoint as a range between contamination levels identified as posing no ecological risk and the lowest contamination levels identified as likely to produce adverse ecological effects."

The NCG wildlife risk assessment is incomplete because it ignores exposures that exceed the NOAEL but are less than the LOAEL, and misses chemical exposures that may result in risk. The use of the NOAELs and LOAELs would change the conclusions of the risk characterization.

NCG should revise the BERA wildlife risk characterization and include comparison of the BERA TDIs to NOAELs in addition to LOAELs.

5. THE USE OF REFERENCE AREA DATA IN THE ECOLOGICAL RISK ASSESSMENT

Recently (February 3, 2017), EPA issued an email explaining their plan for screening the reference area stations. They concluded that utilizing all 8 original reference area selection criteria to screen out sites with high contaminant levels would result in too few stations for a robust comparison. They therefore chose a single criterion (Mean PEC-Q using 17 PAHs) as their screening criteria. EPA states that 0.52 was the highest Mean PEC-Q for the four selected reference areas during the ranking process, but then decided that any station with a Mean PEC-Q above 0.55 would be considered an outlier. Using 0.55 as a cut-off value, 6 stations (4 from Westchester Creek, 1 from Head of Bay, and 1 from Spring Creek) are removed. If 0.52 had been used as a cut-off, an additional 2 sites would have been considered outliers (1 from Gerritsen Creek and another from Westchester Creek). Additionally, EPA directed that comparisons to reference areas be conducted in two ways:

1. Reference Envelope Approach: remove the 6 identified outliers from the analysis
2. Individual Reference Area Comparisons: no removal of outliers; all stations will be used

Questions that arise from this:

1. How was 0.55 chosen? There seems to be a logic step missing from "0.52 was the highest Mean PEC-Q for the four selected reference areas during the ranking process" and "therefore any Mean PEC-Q above 0.55 will be considered an outlier."
2. How will reference areas, specifically values generated using the reference envelope approach, be used? As toxicity test reference areas? Or to calculate background concentrations?
3. How will individual reference area comparisons be interpreted? For example, a single station might be toxic compared to Spring and Gerritsen Creeks but not toxic when compared with Head of Bay or Westchester Creek. Will that station ultimately be deemed toxic or non-toxic? It will be important to determine an analysis methodology up front so it does not appear that methods are being selected after-the-fact in order to select a desired outcome.

The EPA recommendations appear to be exploratory in nature and do not adopt specific methods for comparing site and reference areas or making comparisons among reference areas. EPA should specify a clear and explicit methodology for making such comparisons and provide a clear basis for making decisions based upon the results emanating from the application of these methods.

There are real implications because the stations used as reference will affect (1) what is considered toxic at the site (the cleaner the reference area, the more likely that a site station will be toxic in comparison), and (2) may also affect what EPA considers as a background concentration and therefore what the clean-up level should be.

Discrete Comparisons to Each Reference Area:

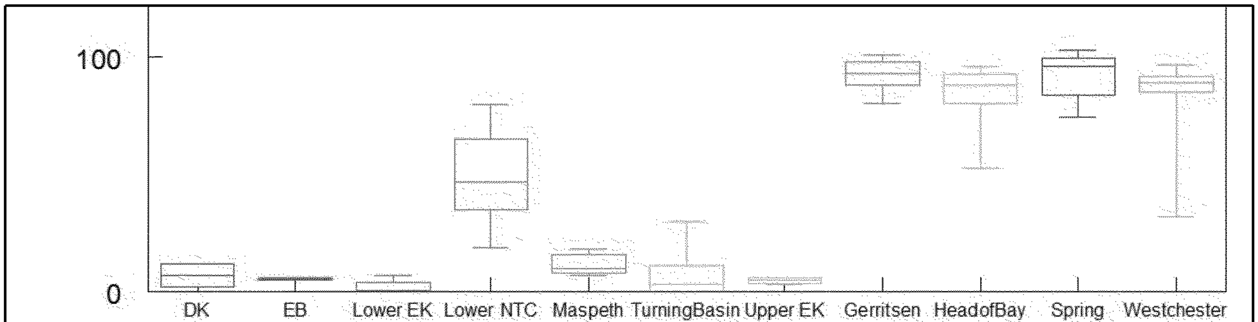
The EPA directs NCG to compare the toxicity and benthic data in the Study Area to each reference area separately. NCG maintains that the work plan requires that the data from all reference areas be lumped. However, the work plan is vague on this issue, and can easily be interpreted to support either approach. Specifically the language in Table 2.2 of the RI Work Plan Volume 1 states that 10- and 28-day toxicity test results should be evaluated through a *"comparison of survival, growth and reproduction of amphipods in Study Area sediments to reference area sediments,"* and that benthic macroinvertebrate metrics should be evaluated through a *"comparison of metrics to reference locations."*

The directions for how to use this information in the ecological risk assessment are vague, and are not clear whether study area data should be compared to each individual reference area separately or all reference area data combined.

Consequences of Individual Reference Area Comparisons

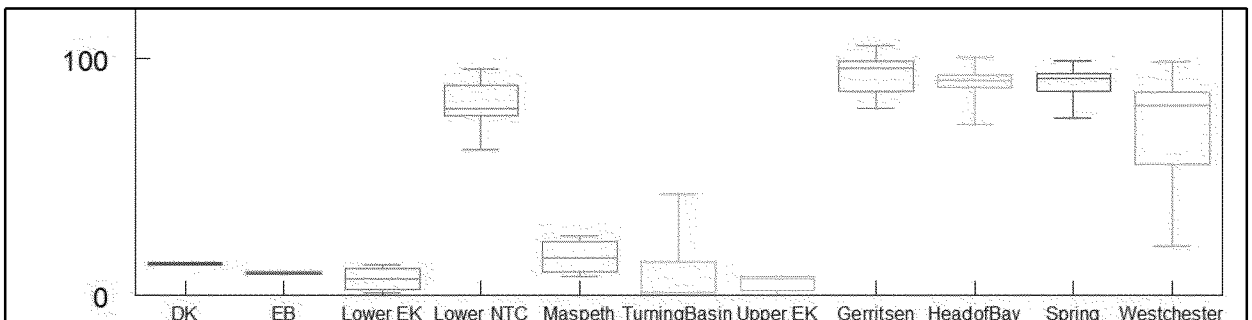
It is not clear how EPA can accomplish their stated goal of using the individual comparisons to clarify the separate contributions of CSOs vs Industrial discharge. In comment ID No. 125, the EPA states that "*the reason four areas were selected that represented four separate categories was to collect data to determine if specific sources of contamination (i.e., industrial discharges and CSO discharges) could be distinguished from each other.*"

The City has previously compared the study area and individual reference area toxicity test results from the NCG program. Figures 5 and 6 show these comparisons for both 10- and 28-day toxicity tests. These figures show how closely the results from all four reference areas are to each other, and how divergent all the study area sites are from any individual reference area. Therefore, whether the site stations are compared to the combined reference areas, or to each reference area separately, the results will be the same. Reference area toxicity data shows that toxicity is not correlated with presence of CSOs (or MS4s, which are also an input at all of these reference areas).



	DK	EB	Lower EK	Lower NTC	Maspeth	TurningBasin	Upper EK	Gerritsen	HeadofBay	Spring	Westchester
SIZE	2	2	4	15	3	5	3	12	12	12	12
5%	2	5	0	18.5	7	0	3	80.5	52.9	74.5	32.2
10%	2	5	0	26	7	0	3	83.5	64.3	77.5	33.4
25%	2	5	0	34.75	7.75	0	3.5	88	80.5	84	85.5
MEDIAN	7	5.5	0.5	47	10	3	5	93	88	96	89
75%	12	6	4	65	16	11.25	5.75	98	92.5	99.5	91.5
90%	12	6	7	70	18	30	6	100.3	95.3	101.6	95.6
95%	12	6	7	79.75	18	30	6	100.9	95.9	102.8	96.8
MEAN	7	5.5	2	48.7333	11.6667	7.6	4.6667	92.3333	84.3333	91.9167	80.6667
STDDEV	7.0711	0.7071	3.3665	18.7711	5.6862	12.7004	1.5275	6.4149	12.8794	9.6903	22.5483
STDERR	5	0.5	1.6833	4.8467	3.283	5.6798	0.8819	1.8518	3.718	2.7973	6.5091
NOTCH	11.1304	1.113	3.1481	12.2943	7.4975	7.9194	2.0448	4.544	5.4528	7.0431	2.7264

Figure 5. NCG 10-day toxicity test survival data: reach-by-reach comparison.



	DK	EB	Lower EK	Lower NTC	Maspeth	TurningBasin	Upper EK	Gerritsen	HeadofBay	Spring	Westchester
SIZE	2	2	4	15	3	5	3	12	12	12	14
5%	13	9	1	61.5	8	0	0	79	71.8	74.9	20.6
10%	13	9	1	66	8	0	0	79	76.6	80.3	43.7
25%	13	9	2.5	76	10	0	1.75	86.5	87.5	86.5	55
MEDIAN	13.5	9.5	7	79	16	1	7	96	90.5	91.5	80.5
75%	14	10	11.5	88.75	22.75	14.5	7.75	99	93	93.5	86
90%	14	10	13	95	25	43	8	101.8	96.8	98.3	93.7
95%	14	10	13	95.75	25	43	8	105.4	100.4	98.9	98.6
MEAN	13.5	9.5	7	81.3333	16.3333	9.8	5	93	89.1667	89.9167	72.2143
STDDEV	0.7071	0.7071	5.4772	10.4994	8.5049	18.6735	4.3589	8.7074	7.7557	6.8684	23.0023
STDERR	0.5	0.5	2.7386	2.7109	4.9103	8.351	2.5166	2.5136	2.2389	1.9827	6.1476
NOTCH	1.113	1.113	7.0833	5.1819	11.5871	10.2072	5.4528	5.68	2.4992	3.1808	13.0414

Figure 6. NCG 28-day toxicity test survival data: reach-by-reach comparison.

6. THE ESTIMATION OF BSAFS SHOULD FOLLOW BURKHARDT'S RECOMMENDATIONS

EPA required that BSAFs be developed for each of the Study Area segments, rather than for the Study Area as a whole. The estimation of BSAFs should follow the recommendations developed by EPA (Burkhardt 2009), which include:

- Estimating the BSAF as the ratio of lipid normal tissue concentrations to TOC normal sediment concentrations;
- Estimating the BSAF by averaging paired measurements of lipid normalized tissue and TOC normalized sediment from areas with similar conditions rather than the use of the slope of a regression line using these parameters; and
- Not combining paired data from areas with highly heterogeneous conditions (as occurs among the various reaches of Newtown Creek).

There are 13 stations (with five replicates per station) with paired polychaete and sediment chemistry data from the bioaccumulation testing (Figure 7). Dutch Kills, East Branch, Maspeth Creek, and Whale Creek have only one station each. English Kills has two stations, the Turning Basin has three stations and lower Newtown Creek has four stations. Therefore, there is no way to estimate variability within a segment for those segments that have only one or two stations. The City recommends that NCG follow the Burkhardt (2009) recommendations and:

- Combine only those stations that have similar conditions (e.g. grain size, TOC, etc.);
- Calculate a BSAF for those areas of similar conditions as the average of the paired data; and,
- Use both lipid normal and TOC normal parameters to make the calculations.

TOC was not measured in the replicate bioaccumulation tests. These calculations should be made using the TOC measured with the bulk sediment chemistry.

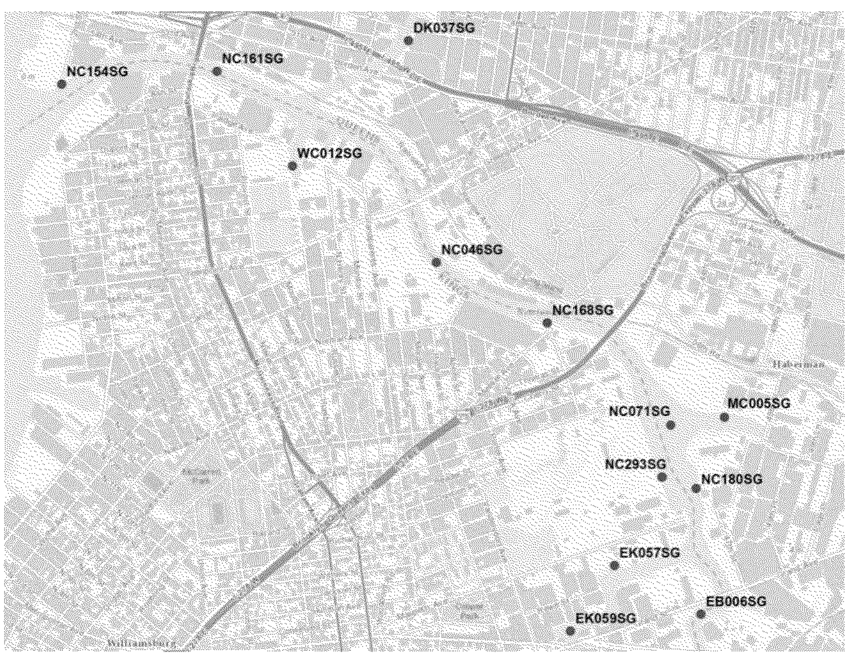


Figure 7. NCG Phase II Bioaccumulation Sampling Sites

References:

Burgess, R.M., W.J. Berry, D.R. Mount, and D.M. Di Toro. 2013. Mechanistic sediment quality guidelines based on contaminant bioavailability: equilibrium partitioning sediment benchmarks. Environmental Toxicology and Chemistry. 32(1):102-114.

Burkhardt, L. 2009. Estimation of Biota Sediment Accumulation Factor (BSAF) From Paired Observations of Chemical Concentrations In Biota And Sediment. U.S. Environmental Protection Agency Office of Research and Development National Health and Environmental Effects Research Laboratory Mid-Continent Ecology Division Duluth, Minnesota, EPA/600/R-06/047, February, 2009.

EPA. 1997. Ecological Risk Assessment Guidance for Superfund: Process for Designing and Conducting Ecological Risk Assessments. Interim Final. Washington, DC: US Environmental Protection Agency, Solid Waste and Emergency Response. EPA 540-R-97-006. June 1997.

EPA. 2003. Procedures for the Derivation of Equilibrium Partitioning Sediment Benchmarks (ESBs) for the Protection of Benthic Organisms: PAH Mixtures. U.S. Environmental Protection Agency, Office of Research and Development. EPA-600-R-02-013. November 2003.

EPA. 2005. Procedures for the Derivation of Equilibrium Partitioning Sediment Benchmarks (ESBs) for the Protection of Benthic Organisms: Metal Mixtures (Cadmium, Copper, Lead, Nickel, Silver, and Zinc). United States Environmental Protection Agency. EPA-600-R-02-011.

EPA. 2014. Memorandum on Toxicity Testing and Ecological Risk Assessment Guidance for Benthic Invertebrates, FROM: Donald J. Brady, Ph.D., Director, Environmental Fate and Effects Division (7507P), Office of Pesticide Programs; TO: Environmental Fate and Effects Division (7507P) Office of Pesticide Programs.

Kennedy, A.J., J.A. Stevens, G.R. Lotufo, J.D. Farrar, M.R. Reiss, R.K. Kropp, J. Doi, and T.S. Bridges. 2009. Comparison of Acute and Chronic Toxicity Methods for Marine Sediments. Marine Environmental Research, Elsevier, 2009, 68 (3), pp.118.

Mount, DR. 2010. Considerations for Predicting the Effects of Petroleum in Sediments. Office of Research and Development National Health and Environmental Effects Laboratory, Mid-Continent Ecology Division, Duluth, MN.

Stanley et al., 2010. Evaluation of reduced volume procedures for acute toxicity tests using the estuarine amphipod, Leptocheirus plumulosus. Env. Tox. Chem. 29(12):2769-2766

USACE and EPA Region 2, 2016. Guidance for Performing Tests on Dredged Material Proposed for Ocean Disposal, Final April 2016